

Twitter: a complementary tool to monitor seasonal influenza epidemic in France ?

Pascal Vilain¹, Luce Menudier², Laurent Filleul²

¹Regional office of French National Public Health Agency, Saint-Denis, Réunion, ²French National Public Health Agency, Saint-Maurice, France

Objective

To investigate whether Twitter data can be used as a proxy for the surveillance of the seasonal influenza epidemic in France and at the regional level.

Introduction

Social media as Twitter are used today by people to disseminate health information but also to share or exchange on their health. Based on this observation, recent studies showed that Twitter data can be used to monitor trends of infectious diseases such as influenza. These studies were mainly carried out in United States where Twitter is very popular [1-4]. In our knowledge, no research has been implemented in France to know whether Twitter data can be a complementary data source to monitor seasonal influenza epidemic.

Methods

For this exploratory study, an R program allowing to the collection, pre-processing (geolocation and classification) and analysis of Tweets related to influenza-like illness was developed.

Collection

Stream API was used to collect Tweets in French language that contained terms “grippe”, “grippal”, “grippaux” without to specify geolocation coordinates.

Pre-process

In order to identify Tweets localized in France, a combination of automated filters has been implemented. At the end, were retained:

- Tweets with geolocation coordinates in France (GPS coordinates, country code, country, place name)
- Tweets whose place indicated in user’s profile matched with a city, department or region of France
- Tweets included FR-related time zone but excluding all Tweets reporting a FR time zone but a non-FR place-code.

In the second time, a support vector machine (SVM) classifier was used to filter out noise from the database. To train the classifier, 1500 Tweets were randomly sampled. Each of these 1500 training Tweets was manually inspected and tagged as valid or invalid according to the likelihood that they indicated influenza-like illness. This hand-tagged training set was converted to vector representation using their term-frequency-inverse document frequency (TF-IDF) scores. These TF-IDF vectors were then input to the SVM for training. To evaluate performances of the classifier: accuracy, recall and F-measure were calculated from a 1000 randomly sampled Tweets manually tagged.

Analysis

Data collected over the period from August 8, 2016 to March 26, 2017 were compared to those of the French syndromic surveillance system SurSaUD® (OSCOUR® and SOS Médecins network) [5] by Spearman’s rank correlation coefficient.



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ethical

In accordance to the National Commission on Informatics and Liberty, information about user account were removed in database except location variables. Usernames contained in the text of the tweet have also been deleted.

Results

Over the study period, the system collected 238,244 influenza-related Tweets of which 130,559 were located in France. After a cleaning step, 22,939 Tweets were classified by the algorithm as an influenza-like illness (ILI). The performances of the classifier were 0.739 for accuracy, 0.725 for recall and 0.732 for F-measure. Figure 1 shows that the weekly number of ILI Tweets follows the same trend as the weekly number of ED visits and physicians consultations for ILI. Regardless of data source, Spearman's correlation coefficients were positive and statistically significant at the national level and for each region of France (Table 1).

Conclusions

This exploratory study allowed to show that Twitter data can be used to monitor the epidemic of seasonal influenza in France and at regional level, in complementarity with existing systems. The system needs to be improved to confirm the trends observed during the next influenza epidemic.

Acknowledgement

We thank Philippe Oesterle and Jean-Bernard Candapanaiken from Regional Health Agency in Indian Ocean, Céline Caserio-Schönemann from French National Public Health Agency and Luc Vitrant. We also thank all physicians of OSCOUR® and SOS Médecins networks.

References

1. Broniatowski DA, Paul MJ, Dredze M. 2013. National and local influenza surveillance through Twitter: An analysis of the 2012-2013 influenza epidemic. *PLoS One*. 8(12), e83672. [PubMed](#)
<https://doi.org/10.1371/journal.pone.0083672>
2. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, et al. 2013. Influenza-like illness surveillance on Twitter through automated learning of naïve language. *PLoS One*. 8(12), e82489. [PubMed](#)
<https://doi.org/10.1371/journal.pone.0082489>
3. Paul MJ, Dredze M, Broniatowski D. 2014. Twitter improves influenza forecasting. *PLoS Curr*. 6. [PubMed](#)
4. Allen C, Tsou MH, Aslam A, Nagel A, Gawron JM. 2016. Applying GIS and machine learning methods to Twitter data for multiscale surveillance of influenza. *PLoS One*. 11(7), e0157734. [PubMed](#)
<https://doi.org/10.1371/journal.pone.0157734>
5. Ruello M, Pelat C, Caserio-Schönemann C, et al. 2017. A regional approach for the influenza surveillance in France. *Online J Public Health Inform*. 9(1), e089. <https://doi.org/10.5210/ojphi.v9i1.7671>



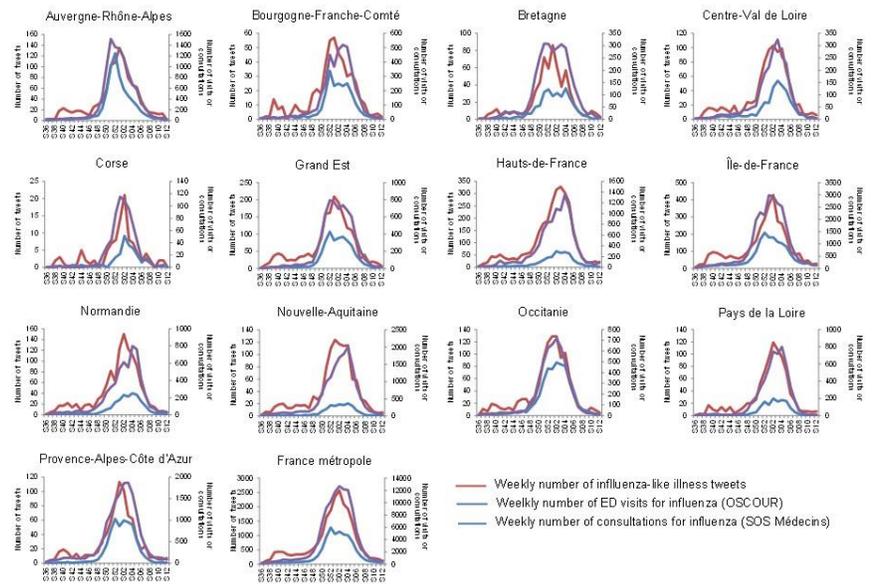


Figure 1. Epidemic curves of weekly number of ILI Tweets and weekly number of visits (OSCOUR®) or consultations (SOS Médecins) for ILI by region of France, W36-2016 to W12-2017

Table 1. Spearman's rank correlation coefficient between ILI Tweets visits (OSCOUR®) or consultations (SOS Médecins) for ILI by region of France, W36-2016 to W12-2017.

Region of France	OSCOUR®		SOS Médecins	
	r_s	p	r_s	p
Auvergne-Rhône-Alpes	0.88	<0.001	0.86	<0.001
Bourgogne-Franche-Comté	0.85	<0.001	0.85	<0.001
Bretagne	0.83	<0.001	0.88	<0.001
Centre-Val de Loire	0.85	<0.001	0.90	<0.001
Corse	0.67	<0.001	0.70	<0.001
Grand-Est	0.87	<0.001	0.84	<0.001
Hauts-de-France	0.85	<0.001	0.88	<0.001
Île-de-France	0.81	<0.001	0.85	<0.001
Normandie	0.83	<0.001	0.87	<0.001
Nouvelle-Aquitaine	0.81	<0.001	0.89	<0.001
Occitanie	0.80	<0.001	0.87	<0.001



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pays de la Loire	0.90	<0.001	0.86	<0.001
Provence-Alpes-Côte d	0.87	<0.001	0.86	<0.001
Metropolitan France	0.89	<0.001	0.90	<0.001



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons AttributionNoncommercial 4.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.